Stratum Boundary Determination (CSRF-E)[i]

Of the various methods available to determine stratum boundaries, New York State has opted to use the *cumulative of the square root of the frequency* method. This method is one that is well documented in many statistics textbooks and one that is well suited for determining stratum boundaries for accounting populations. This method allows for greater efficiency in the sampling process compared with setting the boundaries using judgment only. Using judgment in setting up stratum boundaries is a perfectly acceptable method, but may not be the most efficient in many cases. The department is committed to performing audits using the most efficient and accurate methods available.

Calculation of stratum boundaries using the *cumulative of the square root of the frequency* method
The table below illustrates how stratum boundaries are determined using the *cumulative of the square root of the frequency* method.

Step 1	Evaluate the population and determine if transactions below a certain dollar level should be eliminated from being sampled based on the dollar significance of those transactions. In this example, no transactions under $100 will be reviewed. This is a decision that is made using auditor judgment and has no statistical significance.
Step 2	Evaluate the population and determine what the high dollar cutoff will be for detailed review. In this example (on page 18), all items in the amount of $10,000 and up will be reviewed in detail.
Step 3	Stratify the remaining transactions within the population of items to be sampled ($100 - $9999.99) using a set dollar interval between the strata ($500). We use $500 for the current example to limit the number of strata which need to be displayed. In actuality, $100 intervals are used.
Step 4	Determine the frequency ***f*(y)** for each dollar range. This is nothing more than the number of transactions within each dollar range.
Step 5	Calculate the square root of the frequency for the first sampled range (100 - 499.99), which calculates out to be 146.5. Now calculate the square root for the next sampled range (500 - 999.99), which calculates out to be 94.1. Since we are accumulating this column, each result gets added to the number before it. The cumulative of the first two stratum equals the 146.5 + 94.1 which equals the 240.6. Continue this process for each of the sampled strata.
Step 6	Once the *cumulative square root of the frequency* for the last sampled strata (9500 - 9999.99 in this case) is calculated, the stratum can be determined. For the final result, 770.1 is divided by the number of sampled strata desired (6 in this case, but can vary between 3-9 sampled strata). This result is the test interval of 128.4.

Step 7 The first strata range is determined by comparing the test interval of 128.4 against the calculated cumulative square root of the frequency for each dollar range and selecting the range that is closest to this test interval amount. In the immediate instance, 128.4 is closest to the 146.5 for the first sampled strata (100 - 499.99). The second stratum is determined by taking 2 X 128.4 and finding the closest range. Strata 2 will encompass the range of 500 - 999.99 since 240.6 is the closest interval to 256.8. The third stratum is determined by taking 3 X 128.4 and finding the closest range. Strata 3 will encompass the range of 1000 - 2499.99 since 401 is the closest interval to 385.2. This process will continue until all 6 sampled strata are determined.

Dollar Range (interval)	**Number of items frequency f(y)**	**Cumulative Square root of the frequency** $\sum\sqrt{f(y)}$	**Strata #**	**Strata range**
0 - 99.99	20090			***Eliminated***
100 - 499.99	21472	***146.5***	***1***	***100 - 499.99***
500 - 999.99	8850	***240.6***	***2***	***500 - 999.99***
1000 -1499.99	4149	305.0	***3***	***1000 - 2499.99***
1500 - 1999.99	2637	356.4		
2000 - 2499.99	1990	***401.0***		
2500 - 2999.99	1620	441.2	***4***	***2500 - 3999.99***
3000 - 3499.99	1815	483.8		
3500 - 3999.99	1056	***516.3***		
4000 - 4499.99	879	546.0	***5***	***4000 - 6499.99***
4500 - 4999.99	627	571.0		
5000 - 5499.99	598	595.5		
5500 - 5999.99	520	618.3		
6000 - 6499.99	425	***638.9***		
6500 - 6999.99	471	660.6	***6***	***6500 - 9999.99***
7000 - 7499.99	458	682.0		
7500 - 7999.99	429	702.7		
8000 - 8499.99	358	721.6		
8500 - 8999.99	241	737.2		
9000 - 9499.99	275	753.7		
9500 - 9999.99	269	***770.1***		
10000 and up	234		***7***	***Detail***

The population profile below summarizes the results from using the *square root of the frequency* method to determine the stratum boundaries.

Population profile

Strata #	Strata range	Number of items	Population $'s	Standard deviation
1	0 - 99.99	20,090	738,807	-
2	100 - 499.99	21,472	5,432,668	109.89
3	500 - 999.99	8,850	6,368,355	144.67
4	1000 - 2499.99	8,776	14,041,520	431.59
5	2500 - 3999.99	4,491	14,103,433	411.27
6	4000 - 6499.99	3,049	15,469,958	715.53
7	6500 - 9999.00	2,501	20,025,062	1010.26
8	10000 and up	234	6,319,170	-

[i] Taken from the New York State Department of Taxation and Finance's publication 132; *Computer-Assisted Audits – Guidelines and Procedures for Sales Tax Audits*.